

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-242018

(43)Date of publication of application : 29.08.2003

(51)Int.Cl. G06F 12/00
G06F 13/00

(21)Application number : 2002-036265 (71)Applicant : SONY COMMUNICATION
NETWORK CORP

(22)Date of filing : 14.02.2002 (72)Inventor : FUJII FUMIICHIRO

(54) CACHE METHOD AND CACHE SERVER

(57)Abstract:

PROBLEM TO BE SOLVED: To solve such a problem that comfortable services cannot be received because supplied cache data does not meet clients' environment.
SOLUTION: A retrieving part 22when storing a cache entry 26 in a cache storage part 14obtains the environmental information of the client 10 who has requested content dataand retains it by linking it with the environmental information in the cache entry 26. And alsowhen a request for transmission of the content data is made by the client 10the retrieving part 22 obtains the environmental information of the client 10 and extracts the cache entry 26 suited for the environmental information to provide it to the client 10. An estimation part 24 estimates the environmental information of the client 10 from actual communication records between the client 10 and a server 16.

CLAIMS

[Claim(s)]

[Claim 1]While carrying out cash of the same original data transmitted to a client from a server as different processing data of plurality according to environment information of a clientA caching method choosing one processing data according to said environment information from those processing dataand transmitting to a client.

[Claim 2]A caching method transmitted to a client after processing the original data according to environment information of a client while carrying out cash of the single original data transmitted to a client from a server.

[Claim 3]The caching method according to claim 1 or 2wherein said environment

information of the client is attached to said Request to Send at the time of a Request to Send to a server by a client.

[Claim 4]A cache server comprising:

A cash storage parts store which carries out cash of the original data transmitted to a client from a server as different processing data of plurality according to environment information of a client.

A retrieval part which searches cache data to data which said client requires within said cash storage parts store and chooses said processing data according to said environment information of said client when there is a Request to Send from a client to a server.

The communications department which transmits said selected processing data to said client.

[Claim 5]The cache server according to claim 4 wherein said retrieval part chooses said processing data according to said environment information with reference to said environment information attached to said Request to Send.

[Claim 6]A cache server comprising:

A cash storage parts store which carries out cash of the original data transmitted to a client from a server.

A retrieval part which searches cache data to data which said client requires within said cash storage parts store when there is a Request to Send from a client to a server

A processing section which processes said searched cache data according to said environment information of said client.

The communications department which transmits said processed cache data to said client.

[Claim 7]The cache server according to claim 6 wherein said processing section processes said cache data with reference to said environment information attached to said Request to Send.

[Claim 8]The cache server according to claim 4 or 6 by which an estimating part which presumes said environment information of the client according to a communication track record between a client and a server being included further.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention]This invention relates to the cash art of data. This invention relates to the caching method and cache server which provide a client with the cache

data according to the environment of the client especially.

[0002]

[Description of the Prior Art]The broadband network age can have come and a user can always gain now the high-speed Internet access environment of connection. The service currently kept at arm's length with restriction of transmission speed conventionally [such as distribution of an animation] is also considered that use spreads from now on. Exchanging an E-mail with a photograph using a cellular phone also comes to be performed and a personal digital assistant and a cellular phone also occupy a position important as Internet connectivity apparatus increasingly in addition to the usual personal computer.

[0003]It is necessary under such a situation for a server to adjust the capacity of data, the image and the audio quality to provide according to the gestalt of a user's terminal or a network communication band. Since the throughput of a server or a router, the bandwidth of a communication line etc. have restriction when distributing contents using a network, a cache server is provided on a network and generally holding the contents of a server temporarily is performed in order to secure processing performance.

[0004]

[Problem(s) to be Solved by the Invention]However, in order that a cache server may make the data of the contents by which the Request to Send was carried out correspond to a data name and may carry out cash from a user's terminal, it will provide the same data uniformly to the terminal which required the data of the same name. Therefore, even if the communication performance etc. of the network which the machine performance of a terminal and a terminal have connected are not taken into consideration but it is carrying out cash of the contents of a server with much trouble, the data which does not suit a user's utilizing environment will be sent to a terminal. Therefore, mismatching may arise between users' processing performance and the data supplied and the problem that comfortable service cannot be received may arise.

[0005]This invention is made in view of such a situation and the purpose is in offer of the cash art which can supply cache data suitable for the environment of a user's terminal to a terminal.

[0006]

[Means for Solving the Problem]A mode with this invention is related with a caching method. While this method carries out cash of the same original data transmitted to a client from a server as different processing data of plurality according to environment information of a client, according to said environment information, one processing data is chosen from those processing data and it transmits to a client.

[0007]Another mode of this invention is also related with a caching method. After it processes those original data according to environment information of a client, this method is transmitted to a client while it carries out cash of the single original data

transmitted to a client from a server.

[0008]With environment information of a client a client processes data from a server. They are machine performance of a terminal of a user who is the information concerning system environment of a client which poses a problem in respect of performance when perusing for example is a client's specification of various hardware mounted in the terminal or software, the processing performance of a network which the terminal has connected, etc.

[0009]Also in which above-mentioned caching method, said environment information of the client may be attached to said Request to Send at the time of a Request to Send to a server by a client. Said environment information of the client may be presumed according to a communication track record between a client and a server.

[0010]Another mode of this invention is related with a cache server. This cache server is provided with the following.

A cash storage parts store which carries out cash of the original data transmitted to a client from a server as different processing data of plurality according to environment information of a client.

A retrieval part which searches cache data to data which said client requires within said cash storage parts store and chooses said processing data according to said environment information of said client when there is a Request to Send from a client to a server.

The communications department which transmits said selected processing data to said client.

[0011]Processing data is held as a cache entry in a cash storage parts store. This processing data may not necessarily be restricted as being processed in a cache server but may be beforehand prepared according to environment information of a client in a server. In that case when there is a Request to Send to original data from a client to a server, the cache server can acquire processing data beforehand prepared in a server and can store it in a cash storage parts store. After a cache server acquires original data from a server, it may be processed into two or more processing data according to environment information of a client. Said Request to Send from a client may be attached and provided with environment information, and said retrieval part may choose said processing data according to said environment information with reference to said this attached environment information.

[0012]Another mode of this invention is also related with a cache server. This cache server is provided with the following.

A cash storage parts store which carries out cash of the original data transmitted to a client from a server.

A retrieval part which searches cache data to data which said client requires within said cash storage parts store when there is a Request to Send from a client to a server.

A processing section which processes said searched cache data according to said environment information of said client.

The communications department which transmits said processed cache data to said client.

Said processing section may process said cache data with reference to said environment information attached to said Request to Send.

[0013]Also in which above-mentioned cache server an estimating part which presumes said environment information of the client according to a communication track record between a client and a server may also be included further. Said retrieval part may choose said processing data based on said presumed environment information. Said processing section may process said cache data based on said presumed environment information.

[0014]What expressed arbitrary combination of the above component and this invention as a method a system a computer program a recording medium etc. is effective as a mode of this invention.

[0015]

[Embodiment of the Invention]Drawing 1 shows the composition of the cache system concerning a 1st embodiment. The client 10a acquires the contents data which accessed the server 16 via the Internet and was stored in the contents database 18. The cache server 12 is formed in the node on a network course and holds temporarily the contents data which the client 10a acquires from the server 16 to the cash storage parts store 14. When there is an access request again to the same contents data from the same client 10a or other clients 10b (it is only called the client 10 when naming these generically below) The cache server 12 transmits the contents data by which cash was carried out to the cash storage parts store 14 to the client 10. Thus when there is access to the same contents of the server 16 the cache server 12 can provide the cache data of the contents can reduce concentration of the load to the server 16 and can reduce the traffic between the server 16 and the cache server 12.

[0016]The cache server 12 is put on the place which access such as an access point of the network at the time of being able to provide in two or more network nodes for example seeing from a user and relay points of an upstream network concentrate. The composition of the cache server 12 to be described from now on may be provided in a proxy server as a part of function of a proxy server. The cash storage parts store 14 and the contents database 18 may be formed in the inside of the cache server 12 and the server 16 respectively and may be formed outside as a mass storage device.

[0017]Drawing 2 is a line block diagram of the cache server 12. In hardware these composition is drawing the functional block realized by those cooperation here although it can realize with elements including CPU of a computer and the program etc. which have a cache control facility by software realize. Therefore these

functional blocks are realizable in various forms with the combination of hardware and software.

[0018] The cache server 12 is matched with the environment information of the client 10, carries out cash of the contents data of the server 16, and when there is a Request to Send to contents data from the client 10, it provides the cache data which suit the environment information of the client 10. Therefore, some methods can be considered as follows depending on how the cache server 12 and the server 16 treat the environment information of the client 10 for the environment information of the client 10 by getting to know in a certain form.

[0019] (A1) The method which registers environment information beforehand. The environment information of the client 10 is known beforehand, the environment information is shared between the server 16 and the cache server 12, and the cache server 12 matches and carries out cash of the environment information to send data.

[0020] (A2) the time of the first Request to Send — or the method which registers environment information at the time of updating. When the client 10 requires transmission of contents data of the server 16 for the first time, the client 10 provides environment information in the form attached to the Request to Send. The cache server 12 registers the environment information and associates and carries out cash to send data. The client 10 can update the environment information registered into the Request to Send by attaching new environment information in addition to the time of the first Request to Send also when it was necessary to change the registered environment information.

[0021] (A3) Method which always attaches environment information to a Request to Send. The client 10 attaches environment information to a Request to Send always or as much as possible and requires contents data from the server 16.

[0022] Next, the contents of the environment information of the client 10 are described. Some variations can be considered as follows also to environment information.

[0023] (B1) Version information such as a browser for the contents inspection of the client 10, a viewer of a picture or an animation, and a decoder. With versions of a browser or a viewer, the kind of data which can be processed, data volume, the capacity of cash, etc. change. A decoder is combination of hardware, software, or hardware and software which decodes image data such as MPEG and JPEG. A kind of data volume, etc. of the data which can also process a decoder by a version differ from each other, or the data format which can be processed was decided. Thus, version information becomes one factor when deciding the classification of contents data, a data format, capacity which should be provided to the client 10.

[0024] (B-2) Communication environment information on the client 10. This is the performance information of network, such as bandwidth, etc. of the network which the client 10 has connected.

[0025] (B3) Communication data between the client 10 and the server 16. This is information which opts for the performance, the number, i.e. the hop number, etc. of a

relay node on a communication path etc. of a communication path.

[0026](B4) Hardware information of the client 10. This is the information about hardware abilities of a system such as processing performance of a graphic board or a sound board, CPU performance, resolution of a display and a data transfer rate of a hard disk.

[0027] Thus although it thinks of two or more candidates both the acquisition method of the environment information of the client 10 and the contents of environment information. If the ease of carrying out of realistic realization is taken into consideration it is appropriate to adopt the combination of (B1) and (B-2) as the attachment method of (A3) and contents of environment information as an acquisition method of environment information. That is in the case of the Request to Send to the contents data of the server 16 the client 10 always attaches self environment information to a Request to Send it transmits and the environment information to attach makes it the version information of the client 10 and communication band width. The cache server 12 relates version information and communication band width with send data and holds them to the cache storage parts store 14. When the Request to Send to the server 16 is made from the client 10 When all the identification information of the version information attached to the Request to Send communication band width and the contents data which is demanding transmission is in agreement the cache server 12 transmits cache data instead of the server 16. About the client 10 which performs a Request to Send without attaching environment information same processing is performed supposing the version information and communication band width which were set up beforehand. This method is efficient the following point.

[0028](1) It will always be necessary to hold the correspondence relation between the client 10 and environment information and the management cost of the cache server 12 will increase in the method which transmits environment information only when the client 10 is required. By this method since the client 10 always attaches environment information to a Request to Send it is not necessary to memorize the environment information of each client 10.

[0029](2) It becomes the big load for the client 10 that the client 10 always investigates communication data at the time of a Request to Send and transmits. By this method communication data are not included in environment information since it is only version information and the information of communication band width decided statically what is necessary is to reset up environment information only when the machine environment and communication environment of the client 10 change and load is not applied to the client 10.

[0030](3) The communication band width of the client 10 can be acquired easily and is suitable as a rule of thumb of transmission speed.

[0031](4) A data format is changed by the version of program such as a browser of the client 10 and a decoder in many cases and version information is useful for the

judgment of the data format which can be processed by the client 10.

[0032] The composition and operation of the cache server 12 are explained with reference to drawing 2. The communications department 20 performs data communications between the client 10 and the server 16 via the Internet. The retrieval part 22 is held to the cash storage parts store 14 by making into the cache entry 26 the contents data which had the Request to Send from the client 10 to the server 16. In that case the retrieval part 22 investigates the environment information of the client 10 and associates and holds this environment information in the cache entry 26. The retrieval part 22 searches the data which the client 10 requires from the server 16 within the cash storage parts store 14. In that case the retrieval part 22 investigates the environment information of the client 10, extracts the cache entry 26 whose environment information corresponds, and provides it to the client 10.

[0033] The estimating part 24 presumes the environment information of the client 10 from the communication track record between the client 10 and the server 16. For example, from the information on the time stamp of a packet exchanged between the client 10 and the server 16, the communication band width by the side of the client 10 is presumed. The retrieval part 22 extracts the cache entry 26 which was suitable for the environment of the client 10 using the presumed environment information and provides it to the client 10. Thus, the environment information presumed by the estimating part 24, besides acquiring environment information from the client 10 directly, can be used for the retrieval part 22.

[0034] Drawing 3 is a figure explaining the data structure of the cache entry 26 held at the cash storage parts store 14. The cache entry 26 has the field of the data identification information 30, the environment information 32, and the data 34. The data identification information 30 is information for identifying the contents data by which cash is carried out, and URL which shows the storing position of the contents data as an example is stored. They may be a document name and a file name as other examples. The environment information 32 is environment information of the client 10 which required contents data, for example, values such as the version 6 and 64K bps are stored. The data 34 is cache data of the contents concerned, and that in which size differs from original contents data according to the environment information 32 is stored here.

[0035] Next, the procedure of the cache control by the cache server 12 of the above composition is shown. In drawing 4 (a), version information is 6 and the 1st client 10a is 64K bps in communication band width. The 1st client 10a sends the data transmission request 40 to the cache server 12 (S10). The data transmission requests 40 are "the data A, the version 6 and 64K bps" in this example, including the version information and bandwidth which are the identification information of the contents data which carries out an access request, and the environment information of the 1st client 10a. If the communications department 20 of the cache server 12 receives the data transmission request 40 from the 1st client 10a, the retrieval part 22 will search the cache data of

the data A corresponding to the environment information of the 1st client 10a within the cash storage parts store 14 (S12). In order not to hit into cash the communications department 20 of the cache server 12 transmits the data transmission request 40 to the server 16 (S14).

[0036] In drawing 4 (b) the server 16 receives the data transmission request 40 from the cache server 12. According to two or more environment information three kinds of processing data 50, 52 and 54 is stored in the contents database 18 of the server 16 to the same contents data A. That is they are the processing data 50 corresponding to environment information "version 6 and 64K bps" the processing data 52 corresponding to environment information "version 7 and 64K bps" and the processing data 54 corresponding to environment information "version 7 and 32K bps."

[0037] The server 16 extracts the processing data 50 corresponding to the environment information "version 6 and 64K bps" contained in the data transmission request 40 (S16) and transmits the processing data 50 to the cache server 12 (S18). the retrieval part 22 of the cache server 12 makes a cache entry the processing data 50 received from the server 16 — the cash storage parts store 14 — storing (S20) — the processing data 50 is transmitted to the 1st client 10a via the communications department 20 (S22).

[0038] Next in drawing 5 (a) the 2nd client 10b requires the same contents data A of the server 16. The environment information of the 2nd client 10b is "the version 7 and 32K bps." The 2nd client 10b transmits to the data transmission request 42 at the cache server 12 including this environment information (S24). The cache server 12 searches the cache data of the data A corresponding to the environment information "version 7 and 32K bps" contained in the received data transmission request 42 within the cash storage parts store 14 (S26). In order that the cache data corresponding to the environment information of the 2nd client 10b may not hit the cache server 12 transmits the data transmission request 42 to the server 16 (S28).

[0039] In drawing 5 (b) the server 16 extracts the processing data 54 corresponding to the environment information "version 7 and 32K bps" contained in the data transmission request 42 (S30) and transmits the processing data 54 to the cache server 12 (S32). the retrieval part 22 of the cache server 12 makes the processing data 52 received from the server 16 a new cache entry — the cash storage parts store 14 — storing (S34) — the processing data 54 is transmitted to the client 10 via the communications department 20 (S36).

[0040] Next in drawing 6 (a) the 3rd client 10c requires the same contents data A of the server 16. The environment information of the 3rd client 10c is the same as that of the environment information of the 1st client 10a shown in drawing 4 (a) and is "the version 7 and 32K bps." The 3rd client 10c transmits to the data transmission request 44 at the cache server 12 including this environment information (S38). The cache server 12 searches the cache data of the data A corresponding to the environment information "version 6 and 64K bps" contained in the received data transmission

request 44 within the cash storage parts store 14 (S40).

[0041]In drawing 6 (b)in the cash storage parts store 14the cache data corresponding to the environment information "version 6 and 64K bps" of the 3rd client 10c hitand the retrieval part 22The processing data 50 corresponding to environment information "version 6 and 64K bps" is extracted (S42)and it transmits to the client 10 (S44).

Thuswhen the cache data corresponding to the environment information of the 3rd client 10c hitbetween the cache server 12 and the server 16communication does not occurbut the load to the server 16 is reducedand network traffic is reduced.

[0042]Drawing 7 shows the composition of the cache server 12 concerning a 2nd embodiment. Although two or more cache data were provided to the same contents data in a 1st embodiment corresponding to the environment information of the client 10According to a 2nd embodimentsingle cache data are held to the same contents dataand the cache data are processed according to the environment information of the client 10. Since other composition and operation are the same as a 1st embodimentthey explain only a different portion.

[0043]The cache server 12 has the processing section 28and it processes the cache entry 27 which the retrieval part 22 extracted so that it may be suitable for the environment information of the client 10. When the environment information of the client 10 is not obtained directlylike a 1st embodimentthe estimating part 24 presumes the environment information of the client 10and the processing section 28 processes the cache entry 27 using the presumed environment information.

[0044]Drawing 8 is a data configuration figure of the cache entry 27 stored in the cash storage parts store 14. The cache entry 27 has the field of the data identification information 30 and the data 34and the data 34 is matched with the data identification information 30 in each cache entry 27. The data 34 is original contents data and is the maximum size as data volume. The processing section 28 reduces data volume suitably according to the environment of the client 10or processes this data by changing a compression ratioand provides the client 10 with it.

[0045]As explained aboveeven when the server 16 distributes the contents data of a different mode according to the environment of the client 10 according to the embodiment of the inventionCash of the contents data of a mode which is different also in the cache server 12 corresponding to the server 16 can be carried outand it can provide for the client 10.

[0046]Since it is a method the environment information of the client 10 is limited to version information and communication band widthand the environment information is further attached [a method] for the client 10 to a Request to Send at every Request to SendIn the cache server 12the environment information of the client 10 is efficiently manageable.

[0047]Since the typical environment information defined beforehand is used when the client 10 does not attach environment information to a Request to Sendit can respond also to an exceptional phenomenonand it can prevent so that it may not

interfere with employment of a system.

[0048] In the above this invention was explained based on some embodiments. These embodiments are illustration and it is just going to be understood that modifications various about those each component and combination of each treatment process are possible and that such a modification is also in the range of this invention by the person skilled in the art.

[0049] Although the cache server 12 of a 1st embodiment matches and holds the environment information and cache data of the client 10 as such a modification it may be substantially impossible to obtain the environment information of the client 10. In that case the identification information and cache data of the client 10 are matched and stored in the cache storage parts store 14. When there is a Request to Send from the client 10 the retrieval part 22 investigates the identification information of the client 10 and it may be made to provide the client 10 with the cache data corresponding to the identification information. Since matching and holding cache data to each client 10 has a limit in respect of a storage capacity when the number of the clients 10 with which such environment information is not obtained increases two or more clients 10 without environment information may be treated collectively and it may match with the same cache data.

[0050] In the above-mentioned explanation when the environment information of the client 10 and the environment information of cache data were in agreement the cache data were provided to the client 10 but. As long as there are cache data which environment information approximates it may be made to provide the cache data to the client 10 even if environment information does not carry out full match.

[0051] In the case of the real time communication of the data of a stream system two channels data and control information are used but a sound an animation etc. may include the environment information of a client in this control information.

[0052] Although the size of the data provided according to the environment information of the client 10 was changed in the above-mentioned explanation according to environment information original contents data may be processed by the following methods besides changing data size. In the case of a picture or a sound the compression ratio and compression technology of data may be changed. It is possible to change the quality of a picture or a sound to carry out drop frame if it is an animation or to change resolution if it is a picture etc. When contents are what contains the data of two or more kinds such as a text a picture and a sound like a Web page processing of only a text transmitting and transmitting only a plain text without transmitting layout information without transmitting a picture for example is possible.

[0053] Although it is presumed whether the cache server 12 would acquire the environment information of the client 10 in the above-mentioned explanation machine kind information is acquired from the client 10 and the judgment part which judges the environment information of the client 10 from machine kind information may be

provided further. The cache server 12 may have the correspondence table which matched machine kind information and the specification information of the model may acquire specification information from the acquired machine kind information and may judge environment information such as hardware ability of the client 10.

[0054] Although environment information was attached to the Request to Send in the above-mentioned explanation it may specifically be written in the header unit of the request of HTTP. Although there are some which are called Cookie to a part of function of the browser of the Internet and information can be held to a user's terminal and it can use for it from the server 16 the environment information of the client 10 may be registered into this Cookie. The cache server 12 can acquire environment information with reference to the Cookie of the client 10. When environment information is not set as Cookie a message may be displayed on the client 10 and the input of environment information may be demanded from a user.

[0055] Although the personal computer was assumed as the client 10 in the above-mentioned explanation the clients 10 may be a personal digital assistant and a cellular phone and may communicate with the server 16 by radio.

[0056]

[Effect of the Invention] According to this invention cache data can be provided according to the environment of a client.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is a lineblock diagram of the cache system concerning a 1st embodiment.

[Drawing 2] It is a lineblock diagram of the cache server of drawing 1.

[Drawing 3] It is an explanatory view of the data structure of the cache entry held at the cash storage parts store of drawing 2.

[Drawing 4] It is a figure explaining the procedure of the cache control of a 1st embodiment.

[Drawing 5] It is a figure explaining the procedure of the cache control of a 1st embodiment.

[Drawing 6] It is a figure explaining the procedure of the cache control of a 1st embodiment.

[Drawing 7] It is a lineblock diagram of the cache server concerning a 2nd embodiment.

[Drawing 8] It is an explanatory view of the data structure of the cache entry held at the cash storage parts store of drawing 7.

[Description of Notations]

10 A client and 12 [A contents database and 20 / The communications department 22 retrieval parts and 24 / An estimating part and 26 / A cache entry and 28 / Processing section.] A cache server and 14 A cash storage parts store and 16

A server and 18

1. A server and 18